Multi-scale simulations of biological systems using the OPEP coarse-grained model

Fabio Sterpone, Sébastien Doutreligne, Thanh Thuy Tran, Simone Melchionna, Marc Baaden, Phuong H Nguyen, Philippe Derreumaux

▶ **To cite this version:**

Fabio Sterpone, Sébastien Doutreligne, Thanh Thuy Tran, Simone Melchionna, Marc Baaden, et al.. Multi-scale simulations of biological systems using the OPEP coarse-grained model. Biochemical and Biophysical Research Communications, 2018, 498 (2), pp.296-304. 10.1016/j.bbrc.2017.08.165 . hal-01644564

HAL Id: hal-01644564
https://hal.science/hal-01644564v1

Submitted on 9 Jan 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Multi-scale Simulations of Biological Systems using the OPEP Coarse-grained Model

Fabio Sterpone,[1] Sébastien Doutreligne,[1] Thanh Thuy Tran,[1] Simone Melchionna,[2] Marc Baaden,[1] Phuong H. Nguyen,[1] and Philippe Derreumaux[1,*]

[1]*Laboratoire de Biochimie Théorique, UPR 9080, CNRS Université Denis Diderot, Sorbonne Paris Cité PSL Research University, IBPC, 13 rue Pierre et Marie Curie, 75005, Paris, France*

[2]*Istituto Sistemi Complessi - ISC, Consiglio Nazionale delle Richerche, P.za A. Moro 2, 008185, Italy*

(Dated: July 28, 2017)

Biomolecules are complex machines that are optimized by evolution to properly fulfill or contribute to a variety of biochemical tasks in the cellular environment. Computer simulations based on quantum mechanics and atomistic force fields have been proven to be a powerful microscope for obtaining valuable insights into many biological, physical, and chemical processes. Many interesting phenomena involve, however, a time scale and a number of degrees of freedom, notably if crowding is considered, that cannot be explored at an atomistic resolution. To bridge the gap between reality and simulation, many different advanced computational techniques and coarse-grained (CG) models have been developed. Here, we report some applications of the CG OPEP protein model to amyloid fibril formation, the response of catch bond proteins to two types of fluid flow, and interactive simulations to fold peptides with well-defined 3D structures or with intrinsic disorder.

Keywords: coarse-grained model, amyloid and catch bond protein simulations, hydrodynamics, interactive.

I. INTRODUCTION

Multiscale modeling frameworks have been elaborated for biomolecular systems. They involve either the use of successive simulations of the system with (i) increasing granularities and flexibilities followed by atomistic representations, (ii) on-fly conversion from simplified to all-atom representations or (iii) with different Hamiltonians (e.g., quantum-mechanics and empirical force fields)[1–4].

Many methods for coarse-graining a system have been developed with the aim of studying the physical properties of soluble proteins, transmembrane proteins, lipid bilayers, nucleic acids (DNA/RNA), sugars, surfactants, polymers, sugars, etc. in their ground states [5–26]. The general idea of coarse-graining is to eliminate or average out many degrees of freedom and consists in representing a configuration of the system in terms of beads at lower resolution and developing effective potentials between the CG sites. The reduced number of degrees of freedom and the smoother resulting potential allow for systems of larger sizes to be explored for a much longer time scale. The prize to pay is, however, (i) how to derive effective potentials that maintain the all-atom physical behavior in a water environment in terms of structure, thermodynamics and dynamics[27], and (ii) how to account for the hydrodynamics effects if we use an implicit solvent model[28, 29].

Among the various CG models developed for soluble proteins, the six-bead OPEP protein model, which is basically an all-atom backbone with CG side-chains, except for Proline, and its effective potential have been optimized over the years [30–33]. The main feature of OPEP compared to most other CG force fields is that the formation of backbone H-bonds results from two-body and four-body terms, there are no explicit atomic charges, and ion-bridge potentials between charged amino acids are derived from all-atom potentials of mean force. A flowchart depicting the OPEP force field parametrization scheme can be found in Ref. [27]. Another important feature of OPEP due to the all-atom representation of the peptide bond is that the integration time-step is 2 fs, vs. 10 fs or larger values for MARTINI and other CGs[27]. OPEP has been coupled to many simulation methods: diffusion-controlled Monte Carlo[30, 34, 35], the Activation-Relaxation Technique[36–39], molecular dynamics (MD) and Langevin dynamics[40, 41], replica exchange molecular dynamics (REMD), and Hamiltonian REMD (H-REMD),[27, 42] metadynamics[43], simulated tempering (ST)[44, 45], a greedy approach (PEPFOLD)[46, 47], Lattice-Boltzmann MD[27, 28], and interactive simulations. OPEP has been applied to a wide range of applications. It has been used with success on many soluble proteins of sizes varying between 10 and 200 amino acids[27, 48], and protein complexes[49]. Also OPEP was used as the starting point to generate the flexible coarse-grained DNA and RNA models (HiRe-DNA/RNA)[50, 51].

Of interest is that, by using OPEP, we were the first to observe β-barrels during the assembly of amyloid peptides)[52, 53], that were validated by X-ray micro crystallography and all-atom simulations[54, 55]. We also showed that hydrodynamics (i) change the aggregation mechanisms of Aβ_{16-22} systems consisting of 100 and 1000 peptides[28, 56], and (ii) allow the calculated translational diffusion constants of the 65-residue chy-

*Electronic address: Email:philippe.derreumaux@ibpc.fr

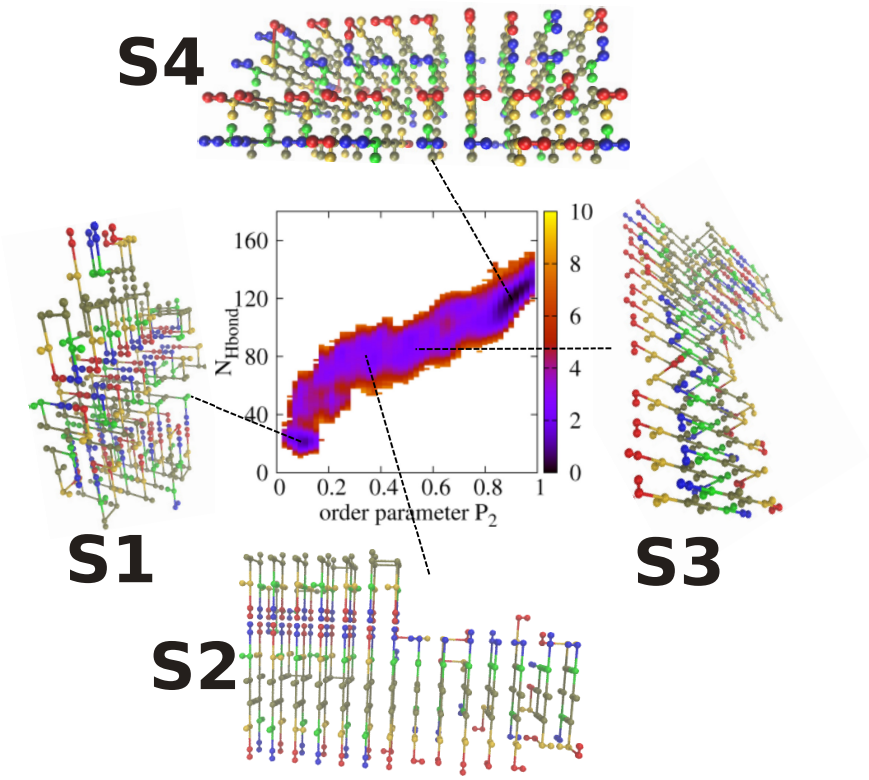

Figure 1: The free energy landscape (in k_BT^*) of the Aβ_{37-42} 40-mers as a function of the order parameter P_2 and total number of inter-chain H-bonds at $T^* = 0.19$ (folding temperature of the monomer). Representative structures corresponding to the states S_i are shown. The color used for the amino acids is G^+ (red) - G (yellow) - V (grey) - V (grey) - I (green) - Ala^- (blue).

motrypsin inhibitor 2 (CI2) protein with crowding conditions varying from 5% to 32% of volume fraction (by tuning the volume of the cubic box containing 70 proteins), to match the NMR observables[29]. In this account, we report new applications of the OPEP protein model to amyloid fibril formation associated with neurodegenerative diseases, catch bond proteins in fluid flow and interactive simulations to fold soluble proteins with well-defined 3D structures or intrinsic disorder.

II. COARSE-GRAINED PROTEIN AGGREGATION LATTICE SIMULATIONS

The basic idea of protein lattice models is that a backbone residue is represented by a bead which is strictly located on the sites of a lattice. A side-chain may also be explicitly included by another bead that represents an amino acid. Different 2D and 3D lattice structures have been proposed, including the simple cubic, face-centered-cubic, hexagonal, triangular and diamond, among others[57]. One of the most studied lattice models is the H-P cubic model, where each bead can be either hydrophobic (H) or hydrophilic (P)[58]. Whereas H-P models only have two residue types, other models with up to twenty residue types have also been considered[59, 60], with interaction energies given by the Miyazawa-Jernigan matrix[61] derived from analysis of the structures of real proteins.

Despite their simplicities, the lattice models have provided useful insights into fundamental questions ranging from protein design, evolution to protein folding and aggregation[62]. For example, they have been used to study the folding pathways[63, 64], the relationships between kinetics, folding rates and thermodynamics[65–67], folding cooperativity[68], the role of the aggregation-prone monomeric conformation and the factors governing fibrillogenesis [69, 70], and the free energy barriers during amyloid fibril formation[71]. Among the models, Frenkel and colleagues developed a sophisticated lattice model including specific side-chain interactions, backbone hydrogen bonds and solvent effects[72–75]. Briefly, a side chain is described by an unit vector d_i, which is not allowed to point in the same direction as the backbone. Thus, only four side-chain directions are allowed for a central residue, and five side-chain directions are possible for the terminus residues. This model has been used to study the fibril formation of a model amyloid peptide with alternating polar and hydrophobic amino acids [74], and the interplay between folding and assembly of a silk-collagen fibril model[76].

Recently, we incorporated the OPEP force field[30, 32,

33, 35, 77] into Frenkel's model and defined a potential energy function of the form[78]

$$E = E_{\rm aa}(\epsilon_{aa}) + E_{\rm hb2}(\epsilon_{hb2}) + E_{\rm hb4}(\epsilon_{hb4}) + E_{\rm steric}(\epsilon_s) + E_{\rm state}(\epsilon_{ss}). \quad (1)$$

Here, $E_{\rm aa}$ represents the total energy of the pairwise interactions ϵ_{aa} between two residues which are separated by a lattice constant a, and their side chain vectors are oriented in the same direction or pointed toward with each other. The interaction energies between 20 amino acids were taken from OPEP parameters[30, 32, 33, 35, 77], and the interaction energies, ϵ_{+-}, between two oppositely charged termini, ϵ_{++} and ϵ_{--} between like-charged termini, ϵ_{+a} and ϵ_{-a} between charged termini and inner residues were determined. The second and third terms are the sums of the two- and four-body hydrogen bond (H-bond) OPEP energies ϵ_{hb2} and ϵ_{hb4}, respectively. A H-bond is formed if two residues are in contact, side chains are oriented in the same direction and both residues are in the strand state. The four-body H-bond term takes the form of sum of weighted products of Gaussian functions each monitoring the existence of an H-bond on the basis of distance criteria. $E_{\rm steric}$ describes the total energy of the steric penalty ϵ_s to prevent clashes between side-chain and backbone atoms. Finally, the secondary structure is modeled by introducing two possible states, coil and strand for each residue, each assigned to an energy value ϵ_{ss}, with $E_{\rm state}$ being the total state energy.

The four-body H-bond term, which is ignored in all current lattice models, plays a crucial role to obtain the correct description of the end-to-end distances and the secondary structures of the Aβ_{16-22} dimer in the lattice representation, as compared to the results obtained from all-atom replica exchange molecular dynamics (REMD) simulations using CHARMM22*/TIP3P and AMBER-f99SB-ILDN/TIP3P force fields[79]. Overall, we found that the calibrated lattice OPEP force field is transferable from the dimer to the trimer of Aβ_{16-22} and also to the dimer and trimer of Aβ_{37-42} by comparing equilibrium structures to those obtained from all-atom REMD simulations.

Using this set of parameters and extensive replica exchange Monte Carlo (REMC) simulations, the model has been employed to determine the critical nucleus size of the experimentally well-characterized Aβ_{16-22} and Aβ_{37-42} peptide segments of the full length Aβ_{1-42} Alzheimer's peptide. At the folding temperature of the monomer, which is defined as the "room-temperature" of the lattice model, we found that more than 90% of the Aβ_{16-22} decamer structures display one 10-stranded β-sheet or two 5-stranded β-sheets which match the microcrystal and solid-state NMR structures of the amyloid fibril, indicating that the nucleus size for fibril formation of Aβ_{16-22} is 10 chains. In contrast, the Aβ_{37-42} decamer is quite disordered with only a 2% population displaying two 5-stranded antiparallel β-sheets and parallel β-strands inside individual sheets[78] matching the fibril structure[80, 81]. Simulations of Aβ_{37-42} 15-mers and 20-mers, starting from disordered states, show that peptides are organized in layers of β-sheets composed of β-strands. The 1-layer architecture is more populated than the 2-layer counterpart and the 3-layer structure is hardly formed. We found, however, that the populations of the perfect fibril state are still low with 2% for the 15-mers and 4% for the 20-mers[82].

In this report, we carried out new REMC simulations for aggregates of 25 and 40 Aβ_{37-42} peptides. A total of 120 replicas covering the temperature range 0.15 - 0.40 (0.19 is the folding temperature of the monomer) were performed for 5 10^{10} MC steps. We found that the population of ordered states with a nematic order parameter $P_2 \geq 0.8$ is 30% for 25-mers and increases to 47% for 40-mers at room temperature. The free energy landscape of the 40-mers at this temperature is shown in Fig.1. The 40-mer exhibits a disordered state S1, representing $\approx$ 22 % of the conformational ensemble with low P_2 values (≤ 0.3) and a small number of intermolecular H-bonds ($N_{\rm H-bond} \leq 30$). Here, the peptides are folded into β-hairpin-like conformations with turns at residues 39-40 or 40-41, and packed in different architectures. The intermediate state S2 state with a population of 17%, ($0.3 \leq P_2 \leq 0.5$, $N_{\rm H-bond} \geq 60$) is characterized by a very complex topology. The first antiparallel two-layer beta-sheet consists of six fully extended antiparallel chains in each layer. The second 2-layer consists of seven less extended parallel chains in each layer. The rest of the peptides are in the folded hairpin-like structures and make end-to-end contacts with the first 2-layers.

In the more ordered intermediate state S3 (population of 14%, $0.5 \leq P_2 \leq 0.7$, $N_{\rm H-bond} \geq 70$), an antiparallel 2-layer beta-sheet structure is formed with 10 fully extended parallel chains in the first layer, and less extended parallel/antiparallel chains in the second layer. The remaining chains are still in the unfolded or beta-hairpin like states, stacked together and sit on the top of the second layer. The topology of this layer is very similar, though not identical to that of the fibril structure[80, 81], thus, it can be considered as a template for fibril growth within the context of a template-assisted self-assembly mechanism. Finally, the FEL displays the almost perfect fibril structure shown in the state S4 (with a population of 47%, $0.8 \leq P_2 \leq 1$, $N_{\rm H-bond} \geq 100$). S4 is formed by four antiparallel 10-stranded β-sheets with parallel β-strands within the individual sheets, except that a few antiparallel chains are still disordered in the two outer layers due to the finite-size effects. The population of nearly 50% of the fibril state suggests that the nucleus size for fibril formation of Aβ_{37-42} is on the order of 40 chains. We are currently using standard Monte Carlo simulations to determine the aggregation mechanisms of the Aβ_{37-42} 40-mers.

III. CATCH-BOND PROTEINS

Important biophysical phenomena involve the response of proteins to mechanical stress generated by the surrounding fluid, also referred to as tensile forces. For example in the vascular system or in the urinary tract, catch-bonds proteins granting cell adhesions are activated by shear flow [83]. Similarly, vessel injuries altering the blood flow activate the response of the multimeric von Willebrand factor (VWF) that favours the anchor of platelets at the injured spots[84].

In this challenging application we have studied the response to fluid flow of a catch-bond protein involved in urinary tract infections by favoring the adhesion of bacterial cells (E. coli) to the host. Piliated E.coli cells attach to target cells via long pili, composed of up to 3,000 copies of the protein FimA. The tip of the pilus is composed of three subunits named FimF, FimG and FimH. The terminal protein FimH (two domains named FimH_L and FimH_P) catches the receptor located at the target cell therefore anchoring E. coli to it. It was proven that the binding to the receptor is stronger under tensile force. It was proposed that the elongation of the multi-domain tip due to fluid flow reduces competitive inter-domain interactions that would cause unbinding [85–87] . A schematic representation of the system is given in the Figure 2.

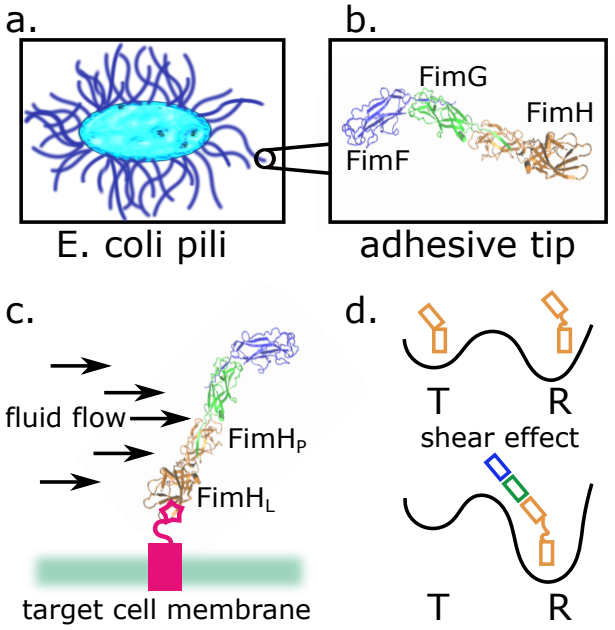

Figure 2: The E. coli's pili (panel a) are essential for bacterial adhesion via tip-receptor interaction. The tip of a pilus is formed by a series of Fim domains (panel b). The interaction with the membrane receptor occurs under physiological fluid flow as depicted in panel c, which intensity strongly favour the high-affinity conformation (or Tense state in allosteric terminology) against the low-affinity conformation (or Relaxed state) as represented in panel d.

Lattice Boltzmann Molecular Dynamics (LBMD) [28] simulations were carried out to investigate the conformational stretch of the three assembled subunits FimF, FimG and FimH that constitute the adhesive tip of the bacterial E.coli pilus, see Figure 2A. The subunits are midsize proteins, composed by 154, 144, 279 amino acids, respectively. Each subunit has been modelled by the OPEP coarse-grained force-field [27, 88] but internal motion was restrained via an elastic network. The three subunits interact with each other via standard OPEP non-bonded interactions. This strategy allowed to increase the integration time-step for the molecular dynamics, $\tau = 10\, fs$ [29], and at the same time to retain the relative subunit conformational flexibility. The system was simulated in a large box of dimensions $L_x = L_y = L_z = 240$ Å. Fluid dynamics was integrated synchronously with the molecular dynamics with a friction coupling $\gamma = 0.1\, fs^{-1}$, a value for the kinetic viscosity matching that of water, and supported on a regular mesh with spacing $dx = dy = dz = 3$ Å[28, 56]. The system was simulated both in absence and presence of fluid flow. For the latter scenario we have considered two kinds of flow perturbations, the shear flow and Poiseuille flow, and scanned various magnitude of the stress. In order to mimic the effect of fluid flow on binding we have linked the binding site of the adhesive domain FimH_L to the edge of the simulation box here representing the target cell membrane. The linking is achieved by freezing the first residue of the FimH subunit. Each simulation was extended up to 1 μs in the case of shear flow and $0.5\mu s$ for Poiseuille flow. To be noted that if transposed to the explicit solvent all-atom representation our system would correspond to about 1.3 million of particles. This size makes clear the need of a multi-scale approach as encoded in the LB technique coupled to an implicit solvent CG model for proteins. Beside the unfeasibility of an all-atom approach, it must also be stressed that the out-of-equilibrium scheme to generate fluid patterns is natural in LBMD, a feature that makes it a perfect tool to model such processes for macromolecules in solution. Other groups have dealt with the problem of shear flow effects on proteins, for instance trying to quantify the minimal shear rate causing functional conformational changes in vWf multi-domain proteins [89], or protein unfolding [90] but all were based on oversimplified representations of the molecular component and used Brownian dynamics that suffer of size scalability for large systems.

We have scanned several shear rates in the range $\dot{\gamma} \in [10^6, 10^9] s^{-1}$. For too high values of shear, $\dot{\gamma} \geq 10^9 s^{-1}$, the three subunits disassemble quite quickly, in the nanoseconds time scales. On the contrary for $\dot{\gamma} \leq 10^8 s^{-1}$ the oligomers remain assembled and respond to the fluid velocity gradient. The response clearly depends on the magnitude of the stress. For the highest tolerable value $\dot{\gamma} = 10^8 s^{-1}$ both the extremities of the protein, the FimH and the FimF units, feel the velocity gradient and bend accordingly, sampling a curved conformation. At the intermediate stress $\dot{\gamma} \sim 10^7 s^{-1}$ the three subunits align along the velocity gradient and stretch.

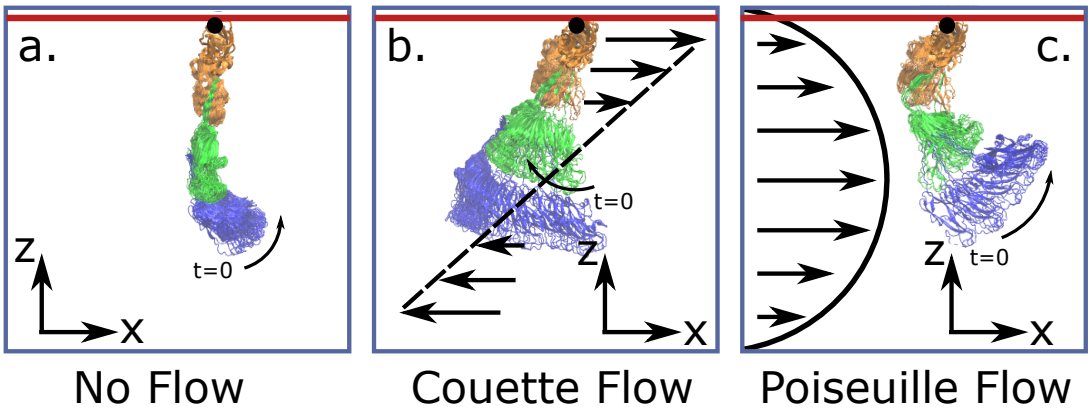

Figure 3: Conformations sampled along the LBMD trajectories. Panel **a** on the left refers to the simulation in absence of fluid flow. The central panel **b** represents the trajectory generated in presence of a Couette flow ($\dot{\gamma} = 10^7 s^{-1}$). The right panel **c** represents the trajectory generated with a Poiseuille flow ($\dot{\gamma} = 10^5 s^{-1}$). The color code for the protein domains is as in Figure 2, the anchoring of the adhesive domain FimH_L to the membrane wall is represented by a black circle.

This regime seems to correspond to what is expected, with the tensile force pushing apart the two domains of the FimH unit. Interestingly, for the lowest sampled shear rate, $\dot{\gamma} \sim 10^6 s^{-1}$, the orientation of the protein chains seems insensitive to the fluid flow, and the conformations acquired resemble that sampled in absence of the fluid flow. An overview of the results is represented in the central panel of Figure 3 where the evolution of the protein conformation during the microsecond trajectory is represented by overlapping representative states separated in time.

Other than shear (Couette) flow, the fluid stress sensed by the FimFGH units could be modeled as a Poiseuille's flow. Again, we have performed a scan of the associated flow rate measured as the derivative of the fluid velocity profile at the wall of the simulation box, $Z = L_z$. For value of the $\dot{\gamma} \geq 10^6 s^{-1}$ the three subunits disassemble, as already observed for high shear rates. At the maximal tolerable value, $\dot{\gamma} \sim 10^5 s^{-1}$, the protein bends according to the velocity parabolic profile, with the FimF unit experiencing the maximal velocity drag, see the Figure 3, right panel. This result shows again that under fluid flow the conformation of the three subunits is dramatically altered and that the chain stretches according to the fluid velocity drag.

In conclusion we have demonstrated that the LBMD machinery is a powerful method in order to model out-of-equilibrium biological processes where the perturbation comes from fluid flow. Our simulations showed that when the appropriate range of shear or flow rate is considered the assembly of the three Fim subunits stretches as effect of the tensile force. According to hypotheses put forward in the literature, as effect of this stretch the binding to the target receptor is reinforced because of conformational changes occurring at the level of the FimH domains. In order to explore this effect, the present work should be extended. Namely, a multi-level representation of the protein should be constructed, with FimH domains modelled by the full flexible OPEP force-field and reserving the elastic network only to the extra units, FimF and FimG. In such away we could explore the inter-domain reorganisation occurring between FimH_L and FimH_P. Moreover, because of complex organisation of the three subunits, the shear and flow directions could be generated also along the Y, or the X-Y directions, so to explore how directionality influences the inter-domain stretching.

IV. INTERACTIVE EXPLORATION OF PROTEIN FOLD SPACE WITH OPEP

We have recently shown that human-guided simulations and standard computational approaches sample RNA conformational space differently[91]. Here we extend our investigations by folding proteins interactively using the OPEP representation. User generated forces enable the system to overcome all energy barriers, and explore very quickly any conformational region of the systems that might be logical to the user based on his knowledge of protein structures, or on the basis of hypotheses, providing therefore an alternative approach to the potential energy surface exploration problem.

Here, we report on interactive simulations of two peptides starting from fully extended states. The first system is the well-studied fragment 41-56 of the B1 immunoglobulin-binding domain of streptoccocal protein G known to form a beta-hairpin in solution and the full protein (PDB entry 1PGB) and has been studied by many standard simulations [27, 36]. The second system is the amyloid-β peptide fragment (Aβ1-34), which in contrast to the two main Aβ1-40 and Aβ1-42 components of amyloid plaques in the brain of patients with Alzheimer's disease, has never been studied experimentally or theoretically [92]. Yet, it has been argued that Aβ1-34 production could contribute to protect against Aβ-mediated toxicity at early stages of the pathology [93].

In order to carry out interactive folding simulations with OPEP, we integrated into the molecular simulation engine interactive molecular dynamics (IMD) features as described in [94–96] with dedicated visualization software allowing the user to apply forces on any specific atoms taken into account by the simulation engine. In our tests, we used UnityMol on the client side [97] with the ability to natively represent coarse-grained OPEP models.

For both peptide systems, we used a pre-defined protocol for the interactive exploration. Starting from full extended states, an initial configuration was obtained after minimization, thermalization and 1 nanosecond of production run at 300 K. The user then applied, from time to time during the second nanosecond interval, forces by means of the computer mouse, inducing an intentional bias on the peptide conformation. Finally, during the 2 - 10 ns interval, the system evolved by means of pure MD simulations. A total of five interactive trajectories (T1 to T5) was performed and differed in the selected atoms subject to the user's forces, and in the orientations of the forces. An additional non-interactive run of 10 nanoseconds was also performed as a control. To illustrate and summarize our interactive exploration in a synthetic manner, we analyzed the generated trajectories by clustering the conformations according to geometric properties. After testing the RMSD as a metric, we settled on ϕ and ψ angles. Details about simulation parameters, software version and the clustering method are given in the supplementary material. Fig. S1 shows a scatter plot of MD-generated configurations along two collective coordinates for each peptide system.

In the interactive simulation of 1PGB41-56, we observed that a configuration with a coil at the N-terminus and a helix spanning Y45-T55 was formed after 1 nanosecond (Fig. S3). We interactively manipulated this structure by pulling the amino acids in the middle of the chain. User forces did not exceed 1211 pN. In the T1 trajectory, the release of the forces led to a β-hairpin, approximately between nanoseconds 1.5 and 2.9, shown as cluster C4 in Figure 4. Afterwards, the peptide folded back to an helix, close to the state we started with, from 4.4 nanoseconds to the end at 10 nanoseconds, corresponding to cluster C8. The mean potential energy along the trajectory is of the order of 10 kcal/mol. These qualitative observations are illustrated in Figure 4. Heavy links between clusters 4 and 8 are mostly explained by the unstable nature of the hairpin with a flexible C-terminus. The same phenomenon is observed in the T4 trajectory where the pulling manipulation was very punctual. This time though, sampling mainly involves clusters 5 and 6.

T2 also reached a β-hairpin secondary structure by using similar manipulations with forces not exceeding 718 pN. This time, however, the hairpin shown as cluster C6 remained stable until the end of the simulation. T2 exhibits a mean potential energy of -10 kcal/mol, much lower than in T1. Similar features are observed in the trajectories T3 and T5.

Analysis of the cluster centers of the five interactive trajectories, each of 10 ns duration, shows that the clusters 1, 2, 3, 4, 5 and 7 share a beta-hairpin arrangement with some variations. Cluster 2 has a more pronounced twist than cluster 1. Clusters 3, 4 and 5 have the same dangling ends and form non-native beta-hairpins, but cluster 5 has a more marked twist. Cluster 7 matches more closely the native experimental beta-hairpin conformation shown in Figure S2. In contrast, clusters 6 and 8 display alpha-helices in the middle of the peptide, spanning residues 46-51 in C6 and 47-53 in C8. As it can be seen from per-cluster potential energy histograms, only clusters 1, 2 and 7 exhibit a mean potential energy below zero. In comparison, the non-interactive control trajectory leads to a full alpha-helix at 10 ns.

We now analyze the result of the simulations for the Aβ1-34 peptide. Without any applied forces, this peptide displays a high number of turns after 1 nanosecond (Figure S4). This feature explains the connection of cluster 1 to all other clusters, as well as the low number of transitions from or to this state. During the T1 trajectory, the peptide attains a structure with two well-formed alpha-helices spanning residues H6-H14 and A21-K28, depicted as cluster C3 in Figure 5.

In T2, the system evolves between two conformational clusters: C6, characterized by an helix between residues E22 and S26 and a beta-hairpin involving the segments E3-R5 and E11-H13, and the cluster C5 with the same hairpin at the N-terminus end and a flexible C-terminus. Analyses of the T3, T4 and T5 trajectories point to the high diversity of structures that can be attained. The cluster C4 is structured as a three-stranded beta-sheet involving the segments V12-Q15, A21-V24 and I32-G33. The clusters C4, C8 and C9 are highly connected, indicating that this double beta-hairpin is a metastable state and easily converts to random coil states (with no secondary structure in C9, and only 15% of alpha-helix in C8). The highly intrinsic disordered property is also demonstrated by the other clusters, with an alpha helix spanning residues H6 to Q15 in C7 and C3, and the C2 cluster free of secondary structure content, and by the similarity of the potential energy histograms between all clusters. Interestingly, the control trajectory of 10 ns leads to a structure reminding an hairpin from D1 to V12 with no hydrogen bond formed yet. This fold is reached after the unfolding of an helix mainly bound by S6:H14 which occured from 6.4 ns to 7 ns, then later between 7.5 ns and 8.5 ns. We find a high propensity of turns from H13 to L34 all along this trajectory.

Overall, the present interactive simulations give interesting insights into the conformational ensembles of both peptides. The peptide 1PGB41-56 has a high propensity to form multiple beta-hairpins and this is consistent with the free energy surfaces obtained by many extensive and computer-demanding simulations.[36]. The formation of an alpha-helix from the non-interactive simulation is not a surprise because the conformational entropy is not taken into account. In addition, a wide exploration of the OPEP 1PGB41-56 energy landscape using

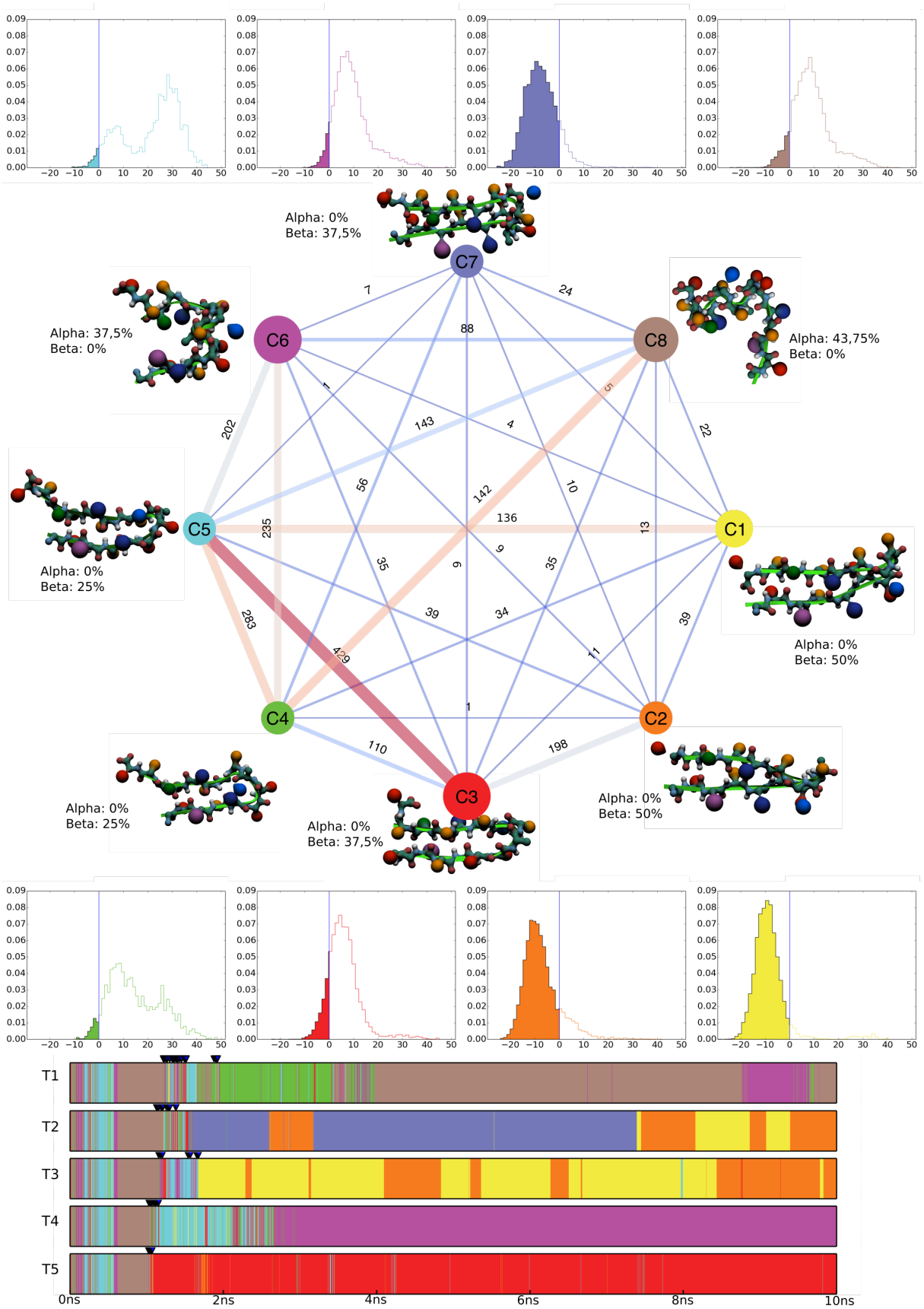

Figure 4: 1PGB 41-56 cluster transition graph. In the large panel, vertices are circles representing clusters colored to match per trajectory barcodes of cluster membership shown on the bottom. For each cluster, we render a representative conformation close to its center with the C-terminus oriented upward. We indicate α-helix and β-sheet ratios obtained from STRIDE assignments performed by VMD and the Timeline plugin visual representation. Vertex circle sizes are proportional to cluster population sizes. Edge widths are proportional to the number of transitions from one cluster to an other that is indicated by a number. The triangles on top of each barcode in the fourth row indicate frames during which the user applied a force. For each cluster, we show the potential energy distribution from all trajectories, with negative energies filled by cluster color.

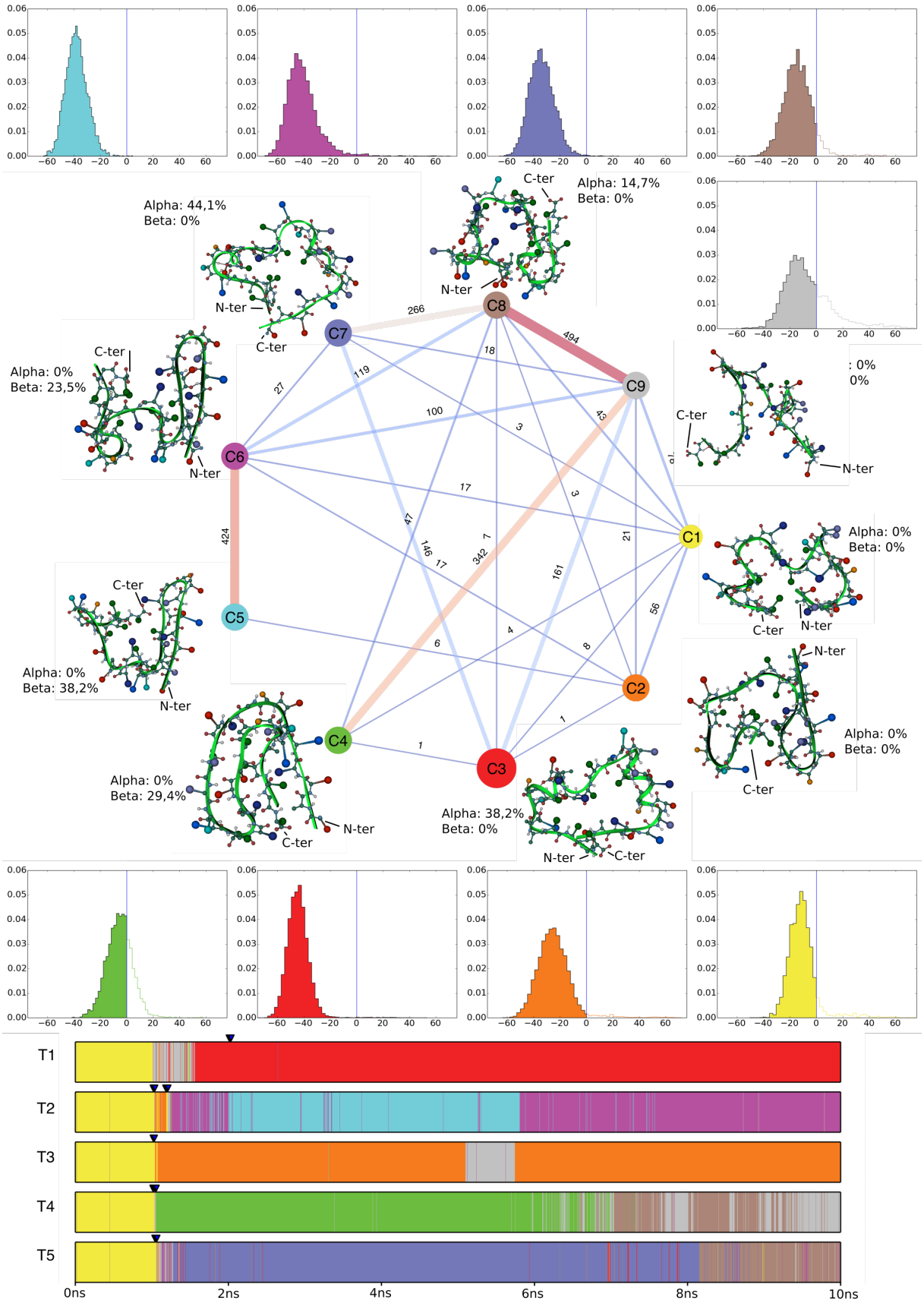

Figure 5: Aβ1-34 cluster transition graph. Same analysis as for 1PGB, except that the N- and C-termini are explicitly shown.

advanced Monte Carlo methods showed a difference of 1 kcal/mol between the minimized energies of the native hairpin and the full alpha-helix [98]. In contrast, the peptide Aβ1-34 is shown very quickly to be highly disordered as expected from the behavior of all amyloid monomeric peptides, giving interest to interactive simulations for testing hypotheses. For 1PGB41-56, each trajectory required around 40 minutes to generate 10 nanoseconds on a high-end workstation against 90 minutes for Aβ 1-34. User interactions with simulations lasted only a couple of minutes. Many simulations were executed in a row. Hence, such an approach represents a fast and easy-to-implement exploration of peptide systems generating a diversity of plausible structures that may guide more extensive and lengthy subsequent runs.

V. CONCLUSIONS AND PERSPECTIVES

We have presented new applications of the OPEP CG model to (i) amyloid fibril formation on a lattice, (ii) the response of catch bond proteins to mechanical stress generated by the surrounding fluid using Lattice Boltzmann MD (LBMD) simulations under Couette and Poiseuille flows, and (iii) the folding of two proteins by interactive simulations guided by humans. While these OPEP simulations provided important insights into three biological systems, we are still improving OPEP-based simulations in three directions.

The first aspect is to optimize the coupling between OPEP and LBMD so as to explore much longer time scales. This can be achieved by a multi-grid approach for the lattice Boltzmann approach, a multiple time step for the integration of the MD equations of motion according to the degree of granularity and flexibility of the system components and porting the code to GPU.

The second aspect is to couple OPEP to a lipid CG model with implicit solvent so as to study the folding and aggregation of proteins at the surface of a membrane or within the membrane, and the interactions of the early formed amyloid-β 40/42 (comprising of 40/42 amino acids) oligomers (e.g. dimers, tetramers, dodecamers), known to be pathological and cytotoxic species in Alzheimer's disease, with well-defined transmembrane protein receptors so as to generate more efficient drugs[99–101].

The third direction is related to improved interactive simulations, starting from plausible models by homology, rather than an a random state, or integrating SAXS, EM and NMR data to provide on-the-fly feedback to the investigator.

VI. ACKNOWLEDGEMENTS

This work was supported by the grants "GRAL", ANR SIMI 12-BS07-0017-01, "ExaViz" (ANR-11-MONU-003) and "DYNAMO" (ANR-11-LABX-0011-01). The authors declare no competing financial interest. FS acknowledges support from the BigChallenge2016 initiative at CINES for generous allocation of computational time.

To whom correspondence should be addressed. E-mail: philippe.derreumaux@ibpc.fr

[1] S. Takada, R. Kanada, C. Tan, T. Terakawa, W. Li, and H. Kenzaki, Acc Chem Res. **48**, 3026 (2015).
[2] H. X. Zhou, Curr Opin Struct Biol. **25**, 67 (2014).
[3] M. Karplus, Angew Chem Int Ed Engl. **53**, 9992 (2014).
[4] W. L. Jorgensen, Cell **155**, 1199 (2013).
[5] M. Levitt and A. Warshel, Nature **253**, 694 (1975).
[6] M. G. Saunders and G. A. Voth, Annu. Rev. Biophys. **42**, 73 (2013).
[7] S. J. Marrink and D. P. Tieleman, Chem. Soc. Rev **42**, 6801 (2013).
[8] P. J. Bond and M. S. P. Sansom, J. Am. Chem. Soc. **128**, 2697 (2006).
[9] D. L. Parton, A. Tek, M. Baaden, and M. S. Sansom, PLoS Comput. Biol. **9**, e1003034 (2013).
[10] A. Y. Shih, A. Arkhipov, P. L. Freddolino, and K. Schulten, J. Phys. Chem. B **110**, 3674 (2006).
[11] R. Devane, W. Shinoda, P. B. Moore, and M. L. Klein, J. Chem. Theory Comput. **5**, 2115 (2009).
[12] A. Leaver-Fay, M. J. O'Meara, M. Tyka, R. Jacak, Y. Song, E. H. Kellogg, J. Thompson, I. W. Davis, R. A. Pache, S. Lyskov, et al., Methods Enzymol. **523**, 109 (2013).
[13] M. Cheon, I. Chang, and C. K. Hall, Proteins **78**, 2950 (2010).
[14] M. Pasi, R. Lavery, and N. Ceres, J. Chem. Theory Comput. **9**, 785 (2013).
[15] M. Zacharias, Proteins **81**, 81 (2013).
[16] A. Davtyan, N. P. Schafer, W. Zheng, C. Clementi, P. G. Wolynes, and G. A. Papoian, J. Phys. Chem. B **116**, 8494 (2012).
[17] N. Basdevant, D. Borgis, and T. Ha-Duong, J. Chem. Theory Comput. **9**, 803 (2013).
[18] T. Bereau and M. Deserno, J. Chem. Phys. **130**, 235106 (2009).
[19] D. Bochicchio and G. M. Pavan, ACS Nano **11**, 1000 (2017).
[20] A. Davtyan, G. A. Voth, and H. C. Andersen, J Chem Phys. **145**, 224107 (2016).
[21] P. A. Netz, R. Potestio, and K. Kremer, J Chem Phys. **145**, 234101 (2016).
[22] M. Invernizzi, O. Valsson, and M. Parrinello, Proc Natl Acad Sci U S A. **114**, 3370 (2017).
[23] W. Zheng and H. Wen, Curr Opin Struct Biol. **42**, 24 (2017).
[24] P. Stadlbauer, L. Mazzanti, T. Cragnolini, D. J. Wales, P. Derreumaux, S. Pasquali, and J. Sponer, J Chem Theory Comput **12**, 6077 (2016).
[25] M. Chavent, A. L. Duncan, and M. S. Sansom, Curr

Opin Struct Biol. **40**, 8 (2016).
[26] S. Kmiecik, D. Gront, M. Kolinski, L. Wieteska, A. E. Dawid, and A. Kolinski, Chem Rev. **116**, 7898 (2016).
[27] F. Sterpone, S. Melchionna, P. Tuffery, S. Pasquali, N. Mousseau, T. Cragnolini, Y. Chebaro, J. F. St-Pierre, M. Kalimeri, A. Barducci, et al., Chem. Soc. Rev. **43**, 4871 (2014).
[28] F. Sterpone, P. Derreumaux, and S. Melchionna, J. Chem. Theory Comput. **11**, 1843 (2015).
[29] M. Chiricotto, F. Sterpone, P. Derreumaux, and S. Melchionna, Philos. Trans. A **374**, 20160225 (2016).
[30] P. Derreumaux, J. Chem. Phys. **106**, 5260 (1997).
[31] P. Derreumaux, J. Chem. Phys. **111**, 2301 (1999).
[32] Y. Chebaro, S. Pasquali, and P. Derreumaux, J. Phys. Chem. B **116**, 8741 (2012).
[33] S. Fabio, P. H. Nguyen, M. Kalimeri, and P. Derreumaux, J. Chem. Theory. Comput. **9**, 4574 (2013).
[34] P. Derreumaux, Phys Rev Lett. **85**, 206 (2000).
[35] F. Forcellino and P. Derreumaux, Proteins **45**, 159 (2001).
[36] G. H. Wei, P. Derreumaux, and N. Mousseau, J. Chem. Phys. **119**, 6403 (2003).
[37] G. H. Wei, N. Mousseau, and P. Derreumaux, J. Chem. Phys. **117**, 11379 (2002).
[38] A. Melquiond, N. Mousseau, and P. Derreumaux, Proteins **65**, 180 (2006).
[39] A. Melquiond, G. Boucher, N. Mousseau, and P. Derreumaux, J. Chem. Phys. **122**, 174904 (2005).
[40] P. Derreumaux and N. Mousseau, J. Chem. Phys. **126**, 025101 (2007).
[41] Y. G. Spill, S. Pasquali, and P. Derreumaux, J Chem Theory Comput. **7**, 1502 (2011).
[42] J. Nasica-Labouze, M. Meli, P. Derreumaux, G. Colombo, and N. Mousseau, PloS Comput Biol. **7**, e1002051 (2011).
[43] A. Barducci, M. Bonomi, and P. Derreumaux, J Chem Theory Comput. **14**, 1928 (2011).
[44] P. H. Nguyen, Y. Okamoto, and P. Derreumaux, J Chem Phys. **138**, 061102 (2013).
[45] T. Zhang, P. H. Nguyen, J. Nasica-Labouze, Y. Mu, and P. Derreumaux, J Phys Chem B. **119**, 6941 (2015).
[46] Y. Shen, J. Maupetit, P. Derreumaux, and P. Tuffery, J Chem Theory Comput. **10**, 4745 (2014).
[47] P. Thevenet, Y. Shen, J. Maupetit, F. Guyon, P. Derreumaux, and P. Tuffery, Nucleic Acids Res. **40(Web Server issue)**, W288 (2012).
[48] M. Kalimeri, P. Derreumaux, and F. Sterpone, J Non Cryst Solids. **407**, 494 (2015).
[49] P. Kynast, P. Derreumaux, and B. Strodel, BMC Biophys. **9**, 4 (2016).
[50] S. Pasquali and P. Derreumaux, J Phys Chem B. **114**, 11957 (2010).
[51] T. Cragnolini, P. Derreumaux, and S. Pasquali, J Phys Chem B. **117**, 8047 (2013).
[52] G. Wei, N. Mousseau, and P. Derreumaux, Biophys. J. **87**, 3648 (2004).
[53] W. Song, G. Wei, N. Mousseau, and P. Derreumaux, J. Phys. Chem. B **112**, 4410 (2008).
[54] A. Laganowsky, C. Liu, M. Sawaya, J. Whitelegge, J. Park, M. Zhao, A. Pensalfini, A. Soriaga, M. Landau, P. Teng, et al., Science **335**, 1228 (2012).
[55] A. D. Simone and P. Derreumaux, J. Chem. Phys. **132**, 165103 (2010).
[56] M. Chiricotto, S. Melchionna, P. Derreumaux, and F. Sterpone, J. Chem. Phys. **145**, 035102 (2016).
[57] W. E. Hart and A. Newman, Handbook of Molecular Biology. pp. 1–24 (2006).
[58] K. A. Dill, Biochemistry **24**, 1501 (1985).
[59] V. I. Abkevich, A. Gutin, and E. I. Shakhnovich, Biochemistry **49**, 11 (1994).
[60] K. B. Zeldovich, I. N. Berezovsky, and E. I. Shakhnovich, J. Mol. Biol. **357**, 1335 (2006).
[61] S. Miyazawa and R. L. Jernigan, Macromolecules **18**, 534 (1985).
[62] S. Kmiecik, D. Gront, M. Kolinski, L. Wieteska, A. E. Dawid, and A. Kolinski, Chem. Rev. **116**, 7898 (2016).
[63] J. Skolnick and A. Kolinski, Science **250**, 1121 (1990).
[64] V. S. Pande and D. S. Rokhsar, Proc Natl Acad Sci U S A. **96**, 1273 (1996).
[65] A. Sali, E. I. Shakhnovich, and M. Karplus, Nature **369**, 248 (1994).
[66] C. J. Camacho and D. Thirumalai, Proc Natl Acad Sci U S A. **90**, 6369 (1993).
[67] R. Melin, H. Li, N. Wingreen, and C. Tang, J Chem Phys. **110**, 1252 (1999).
[68] H. S. Chan, Z. Zhang, S. Wallin, and Z. Liu, Annu Rev Phys Chem. **62**, 301 (2011).
[69] M. S. Li, D. K. Klimov, J. E. Straub, and D. Thirumalai, J. Chem. Phys. **129**, 175101 (2008).
[70] M. S. Li, N. T. Co, C. K. H. G. Reddy, J. Straub, and D. Thirumalai, Phys. Rev. Lett. **105**, 218101 (2010).
[71] A. Irback, S. E. Jonsson, N. Linnemann, B. Linse, and S. Wallin, Phys. Rev. Lett. **110**, 058101 (2013).
[72] S. Abeln and D. Frenkel, PloS Comput. Biol. **4**, e1000241 (2008).
[73] S. Abeln and D. Frenkel, Biophy J. **100**, 693 (2011).
[74] S. Abeln, M. Vendruscolo, C. Dobson, and D. Frenkel, PloS ONE **9**, e85185 (2014).
[75] E. van Dijk, A. Hoogeveen, and S. Abeln, PloS Comput. Biol. **11**, e1004277 (2015).
[76] R. Ni, S. A. M. Schor, M. A. C. Stuart, and P. Bolhuis, Phys. Rev. Lett. **111**, 058101 (2003).
[77] P. Derreumaux, J. Chem. Phys. **107**, 1941 (1997).
[78] T. T. Tran, P. H. Nguyen, and P. Derreumaux, J. Chem. Phys. **144**, 205103 (2016).
[79] P. H. Nguyen, M. S. Li, and P. Derreumaux, Phys. Chem. Chem. Phys. **13**, 9778 (2011).
[80] M. R. S. andS. Sambashivan, R. Nelson, M. I. Ivanova, S. A. Sievers, M. I. Apostol, M. J. Thompson, M. Balbirnie, J. J. W. Wiltzius, and H. T. M. et al., Nature **447**, 453 (2005).
[81] R. Nelson and D. Eisenberg, Curr. Opin. Struct. Biol. **16**, 260 (2006).
[82] M. Chiricotto, T. T. Tranh, P. H. Nguyen, S. Melchionna, F. Sterpone, and P. Derreumaux, Israel Journal of Chemistry p. 10.1002/ijch.201600048 (2016).
[83] W. E. Thomas, V. Vogel, and E. Sokurenko, Ann. Rev. Biophys. **37**, 399 (2008).
[84] T. A. Springer, Blood **124**, 1412 (2014).
[85] O. Yakovenkoa, V. Tchesnokovab, E. V. Sokurenkob, and W. E. Thomas, Proc. Natl. Acad. Sci. USA **112**, 9884 (2015).
[86] M. M. Sauer, R. P. Jakob, J. Eras, S. Baday, D. Eris, G. Navarra, S. Berneche, B. Ernst, T. Maier, and R. Glockshuber, Nature Comm. **7**, 10738 (2015).
[87] V. Kalas, J. S. Pinkner, T. J. Hannan, M. E. Hibbing, K. W. Dodson, A. S. Holehouse, H. Zhang, N. H. Tolia, M. L. Gross, R. V. Pappu, et al., Science Advances **3**,

e1601944 (2017).
[88] F. Sterpone, P. H. Nguyen, M. Kalimeri, and P. Derreumaux, J. Chem. Theory. Comput. **9**, 4574 (2013).
[89] C. E. Sing and A. Alexander-Katz, Biophys. J. **98**, L35 (2010).
[90] P. Szymczak and M. Cieplak, J. Chem. Phys. **127**, 155106 (2007).
[91] L. Mazzanti, S. Doutreligne, C. Gageat, P. Derreumaux, A. Taly, M. Baaden, and S. Pasquali, Biophysical Journal **113**, 302 (2017), ISSN 0006-3495.
[92] J. Nasica-Labouze, P. H. Nguyen, F. Sterpone, O. Berthoumieu, N.-V. Buchete, S. Coté, A. De Simone, A. J. Doig, P. Faller, A. Garcia, et al., Chemical Reviews **115**, 3518 (2015), ISSN 0009-2665.
[93] C. Caillava, S. Ranaldi, I. Lauritzen, C. Bauer, J. Fareh, J.-D. Abraham, and F. Checler, Neurobiology of Aging **35**, 1570 (2014), ISSN 0197-4580.
[94] J. E. Stone, J. Gullingsrud, and K. Schulten, in *Proceedings of the 2001 Symposium on Interactive 3D Graphics* (ACM, New York, NY, USA, 2001), I3D '01, pp. 191–194, ISBN 978-1-58113-292-2.
[95] O. Delalande et al., N. Férey, G. Grasseau, and M. Baaden, J. Comput. Chem. **30**, 2375 (2009), ISSN 1096-987X.
[96] M. Dreher, M. Piuzzi, A. Turki, M. Chavent, M. Baaden, N. Férey, S. Limet, B. Raffin, and S. Robert, Procedia Computer Science **18**, 20 (2013), ISSN 1877-0509.
[97] Z. Lv, A. Tek, F. Da Silva, C. Empereur-mot, M. Chavent, and M. Baaden, PloS one **8**, e57990 (2013), ISSN 1932-6203.
[98] T. Cragnolini, K. H. Sutherland-Cash, D. Wales, S. Pasquali, and P. Derreumaux, Biophysical Journal **106**, 256a (2014), ISSN 0006-3495.
[99] P. H. Nguyen and P. Derreumaux, Acc Chem Res. **47**, 603 (2014).
[100] A. J. Doig and P. Derreumaux, Curr Opin Struct Biol. **30**, 50 (2015).
[101] A. J. Doig, M. P. Del Castillo-Frias, O. Berthoumieu, B. Tarus, J. Nasica-Labouze, F. Sterpone, P. H. Nguyen, N. M. Hooper, P. Faller, and Derreumaux, ACS Chem Neurosci. **8**, 1435 (2017).